

Compare CNN and LSTM Model to Forecast the Stock Price

Tong Wang^{1,a}, Panji Wang^{1,b}, Menghan Yu^{1,c,*}

¹College of Science and Technology, Wenzhou-Kean University, Wenzhou, Zhejiang Province, 325006, China

^awangto@kean.edu, ^bwangpa@kean.edu, ^cyume@kean.edu

*Corresponding author

Keywords: Convolutional Neural Networks, Long Short-Term Memory, Stock Price Prediction

Abstract: The prediction of stock price has been one of the hot research fields in recent years. Although the trend of stock price is non-linear and complex, and affected by the global situation, corporate decisions and other factors, it is difficult to accurately predict the stock price. However, since the change of stock price directly affects the interests of investors, stock price prediction can help people to provide a reference information for future decision-making. In previous studies, neural networks and deep learning have shown good results in stock price prediction. In this paper, two models of LSTM and CNN are used to forecast MRF stock price respectively, and the predicted results are analyzed and compared. We used stock data from January 1, 2013 to May 18, 2018, including four characteristic values: the highest price, the lowest price, the opening price and the closing price. The data of the first 1060 trading days were used to train the two models respectively, and the data of the last 265 trading days were used to test the models. The results show that the prediction effect of CNN model is more accurate than that of LSTM model.

1. Introduction

With the improvement of the world economic system and security system, the stock market is becoming more and more popular, and the number of investors is increasing every year, so how to effectively forecast stock prices has become a popular area of research. The task of stock trend prediction can be seen as a classification task, which usually involves predicting whether prices will rise or fall at the next time step based on past data [1]. The stock market is a complex and dynamic system with a lot of random noise, making it arduous to accurately predict stock movements. Although stock prices do not monitor a random wandering process, predicting the future trend of stocks has proven to be a challenging problem. There are many factors (e.g., company news, industry performance, investor sentiment, and economic factors) that can affect stock price movements [2]. Nevertheless, stock trend forecasting has been of interest to many investors and researchers because successful forecasts can generate huge profits. Fundamental analysis and technical analysis are the two basic forecasting methods used in stock movement forecasting tasks. Fundamental analysis attempts to measure the intrinsic value of a stock by analyzing relevant economic, financial, and other

qualitative and quantitative factors. Technical analysis tends to use historical market data to predict future movements. Often stock movement forecasting tasks are viewed as time series forecasting tasks. Several methods used for time series forecasting can be broadly classified into two categories: traditional linear forecasting methods and non-linear forecasting methods. Linear forecasting methods include autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) [3]. However, financial time series data are nonlinear, which makes it problematic for traditional linear forecasting methods to predict accurately. Recently, many fields have widely expanded machine learning and deep learning techniques. Their application in financial time series forecasting tasks has attracted increasing attention due to their nonlinear mapping and generalization capabilities. A large body of work uses deep learning methods such as CNN, LSTM, and Transformer to achieve respectable results.

Over the past few years, deep learning has excelled in solving many problems such as visual recognition, speech recognition, and natural language processing [4]. As a result, some people have started to apply neural networks to the financial field to achieve results beyond traditional machine learning methods, and the results of real-world strategies from various institutions in recent years have not disappointed, with the addition of reasonable marking and massive factor construction, neural network-based strategies have achieved notable results. Among the distinctive types of neural networks, convolutional neural networks are the most intensively studied [5]. In the early days, it was challenging to train high-performance convolutional neural networks without overfitting due to the lack of training data and computational power. Labeled data and the recent development of GPUs have allowed convolutional neural network research to emerge and achieve first-class results [6]. In this paper, we will look at the recent development of convolutional neural networks and describe the application of time series neural networks LSTM and CNN in the field of stock prediction. In the first half of this paper, we will first analyze the feasibility of LSTM and CNN models, and then we will explain how to use LSTM and CNN models to predict the stock price of Madras Rubber Factory (MRF), an NSE-listed company. Finally, it is concluded from the experimental analysis that the CNN model has a smaller MAPE value compared to the LSTM model and has higher precision and accuracy in stock prediction. Therefore, compared to the LSTM model in predicting the stock price of Madras Rubber Factory, the CNN model would be more suitable for predicting the stock price of this company.

2. Literature review

In recent years, neural networks have achieved downscaling against traditional machine learning methods in both CV and NLP, so some people have started to apply neural networks to finance to achieve results beyond traditional machine learning methods [7]. Deep learning factor stock selection model based on convolutional neural network CNN. Convolutional Neural Network (CNN) is one of the most influential models in the field of computer vision research and application [8]. Similarly, CNN's can produce surprising results for time series processing if time is considered as a spatial dimension, like the height or width of a two-dimensional image. Convolutional neural networks (CNNs) are a common deep learning architecture inspired by living things' natural visual cognitive mechanisms. In 1959, Hubel & Wiesel [9] discovered that animal visual cortex cells are responsible for detecting optical signals. Inspired by this, Kunihiko Fukushima [10] proposed the recognition, the predecessor of CNN, in 1980. In the 1990s, LeCun et al. [11] and others published papers that established the modern architecture of CNN, which was later refined. They designed a multilayer artificial neural network, named LeNet-5, to classify handwritten numbers. Like other neural networks, LeNet-5 can be trained using the Backpropagation algorithm. The CNN was able to derive a valid representation of the original image, which allowed the CNN to identify visual patterns

directly from the original pixels with minimal preprocessing. However, due to the lack of large-scale training data at that time and the computer's computational power to keep up, the results of LeNet-5 for complex problems were not satisfactory. Since 2006, many methods have been devised to overcome the difficulty of training deep CNNs. Among them, the most famous is Krizhevsky et al. [5] proposed a classical CNN structure and made a breakthrough in image recognition tasks. The overall framework of their approach is -called AlexNet, which is similar to LeNet-5 but a bit deeper. after the success of AlexNet, researchers have proposed other refinements, among which the four most famous ones are ZDNet, VGGNet, GoogleNet, and ResNet [12]. In terms of structure, one of the directions of CNN development is to become more layers, with ResNet, the ILSVRC 2015 winner, being more than 20 times larger than AlexNet and more than 8 times larger than VGGNet [13]. By increasing the depth, the network can use the increased nonlinearity to derive an approximate structure of the objective function and at the same time derive a better characterization of the properties. However, doing so also increases the overall complexity of the network, making it demanding to optimize and prone to overfitting. Next, recurrent neural networks are introduced. Recurrent neural networks are a class of neural networks that take sequence data as input, recursively in the direction of sequence evolution and all nodes are connected in a chain-like manner. The research of recurrent neural networks started in the 1980s and 1990s and developed as one of the deep learning algorithms in the early 21st century, among which BidirectionalRNNBi-RNN and Long Short-Term Memory networks (LSTM) are the common recurrent neural networks. Recurrent neural networks have advantages in learning nonlinear features of sequences because of their memorability, parameter sharing, and Turing completeness. Recurrent neural networks have applications in Natural Language Processing (NLP), such as speech recognition, language modeling, and machine translation, and are also used for various types of time series prediction. Long Short-Term Memory is a temporal recurrent neural network, and the paper was first published in 1997 [6]. Due to its unique design structure, LSTM is suitable for processing and predicting important events with long intervals and time series delays. LSTM usually performs better than temporal recurrent neural networks and Hidden Markov Models (HMM), for example, for non-segmented continuous handwriting recognition [14]. In 2009, an artificial neural network model built with LSTM won the ICDAR handwriting recognition competition [15]. LSTM can partially solve the problem that RNNs are prone to gradient disappearance on long sequence problems. The research in this paper focuses on two basic models based on CNN and LSTM. The most common drawback of most existing methods for forecasting stock prices is the inability to remember irrelevant information from the past, to remember important information in the current state, and to predict stock prices when the inability to effectively handle the stochasticity in financial time series data. Our current work attempts to address this problem by explaining the high learning capability of convolutional neural networks (CNNs) and long- and short-term memory (LSTM) networks.

3. Method

3.1. CNN

As one of the representative algorithms of deep learning, CNN has been widely used in various computer vision tasks. And it can also be used in time series prediction, through its function to reduce the number of parameters, improve the efficiency of model learning. The composition structure of CNN includes:

(1)Input layer: Initial processing of row data.

(2)Convolutional Layer: By setting the depth, stride and zero-padding, the convolution kernel is used to calculate the eigenvalues of all regions in the input image, and the feature map storing the eigenvalues of each region of the input image is generated.

(3)Relu Layer: When the input value is less than 0, the output of the activation function RELU is 0, that is, the activation function fails. When the input value is greater than 0, the output is the input value itself. The RELU function is as follows:

$$f(x) = \max(0, x) \quad (1)$$

(4)Pooling Layer: The input data is compressed to reduce the complexity of calculation, reduce overfitting, and help extract the main features. Pooling is divided into two operations: maximum pooling and mean pooling. Generally, a 2*2 matrix is used to scan the data output by the convolutional layer. The maximum value in each region is used to represent the maximum pooling of the eigenvalues of the region, and the mean value of each region is used to represent the average pooling of the eigenvalues of the region.

(5)FC Layer: In the same way as the neurons of traditional neural network are connected, all neurons between each two layers are connected by weights to send the output value to the classifier.

3.2. LSTM

LSTM model was proposed by Hochreiter and Schmidhuber in 1997, which is a special type of RNN [1Achyut Ghosh]. With the ability of memory, LSTM can solve the long-term dependence problem of RNN and obtain a more accurate prediction [1]. Therefore, LSTM is often used in stock price prediction in recent years. The key to the LSTM model is the cell morphology [1], which consists of three parts: the forgetting gate, the input gate, and the output gate. The calculation is as follows:

(1)Input the output value of the previous unit and the input value of the current time into the forgetting gate, and decide which information is allowed to be retained and which is abandoned after calculation. Such as the following formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

where f_t indicates that the output ranges from 0 to 1, 0 indicates that the information is abandoned, 1 indicates that the information is retained, h_{t-1} represents the output of the previous unit, x_t represents the input data at the current time, W_f represents the weight of the forgetting gate, and b_f represents the deviation of the forgetting gate.

(2)Then decide what new information can be retained. There are two parts here. The first part is the sigmoid layer, also known as the input gate, which determines which newly entered information can be updated. The second part is that the tanh layer is designed to generate new information, where a new vector of candidate values is created. The formula is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \sigma(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

where i_t is a control signal that ranges from 0 to 1, W_i represents the weight of the input gate and b_i represents the deviation of the input gate, W_C represents the weight of the candidate input gate and b_f represents the deviation of the candidate input gate.

(3)Renew the state of old cells:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

where the old state is multiplied by f_t to determine the discarded information, and $i_t * \tilde{C}_t$ is the new candidate value, which changes with the updated state.

(4)Determine the output value. The sigmoid layer determines which cell morphology is exported, and then it is processed by tanh layer and exported. The formula is shown below:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

where W_o represents the weight of the output gate and b_o represents the deviation of the output gate.

4. Analysis

The experiment compares CNN and LSTM model in stock price prediction. In order to compare the two methods, we analysed the same training set and test set, and predicted the stock price according to the four main influencing factors: the opening price, the closing price, the highest price and the lowest price.

4.1. Datasets

The stock selected in this experiment is Madras Rubber Factory (MRF), and the data is 1325 daily trading data from January 1, 2013 to May 18, 2018. The daily trading contains four variables – open, close, high and low. The data of the first 1060 trading days are used as the training set, and the data of the last 265 trading days are used as the test set. Some data are shown in Table 1.

Table 1: Partial sample data

Date	Open	High	Low	Close
2013-01-01	12927.0	13380.00	12879.85	13243.25
2013-01-02	13312.4	13435.00	13256.00	13316.20
2013-01-03	13351.0	13365.95	13124.15	13279.10
2013-01-04	13240.0	13418.40	13221.30	13361.10
2013-01-07	13375.0	13439.95	13265.00	13288.80

4.2. Criteria for Evaluation

The root mean square error (RMSE), mean absolute error (MAE) and main absolute percentage error (MAPE) are used to analyse the error of the forecast, and the effect of LSTM and CNN models on the stock price forecast was compared.

The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (8)$$

where reflects the extent to which the measured data deviates from the true value, where \hat{y}_i represents the predicted value and y_i represents the true value. When the RMSE is smaller, the measurement accuracy is higher.

The formula for MAE is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (9)$$

where \hat{y}_i represents the predicted value and y_i represents the true value. When the MAE is smaller, the measurement accuracy is higher.

The formula for MAPE is as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (10)$$

where \hat{y}_i represents the predicted value and y_i represents the true value. When the MAE is smaller, the measurement accuracy is higher.

For the above three error analyses, when their values are 0, the model is perfect and can be predicted 100%.

4.3. Model Implementation

It can be seen from Figure 1 that the validation set loss in LSTM basically tends to 0 after 15 iterations, while CNN is close to 0 in the fifth iteration. In CNN, the verification set error reached the minimum at the 5th iteration, while LSTM reached the minimum at the 15th iteration. In a comprehensive view of the two models, CNN performs better in training data, and the loss and error generated are smaller than LSTM.

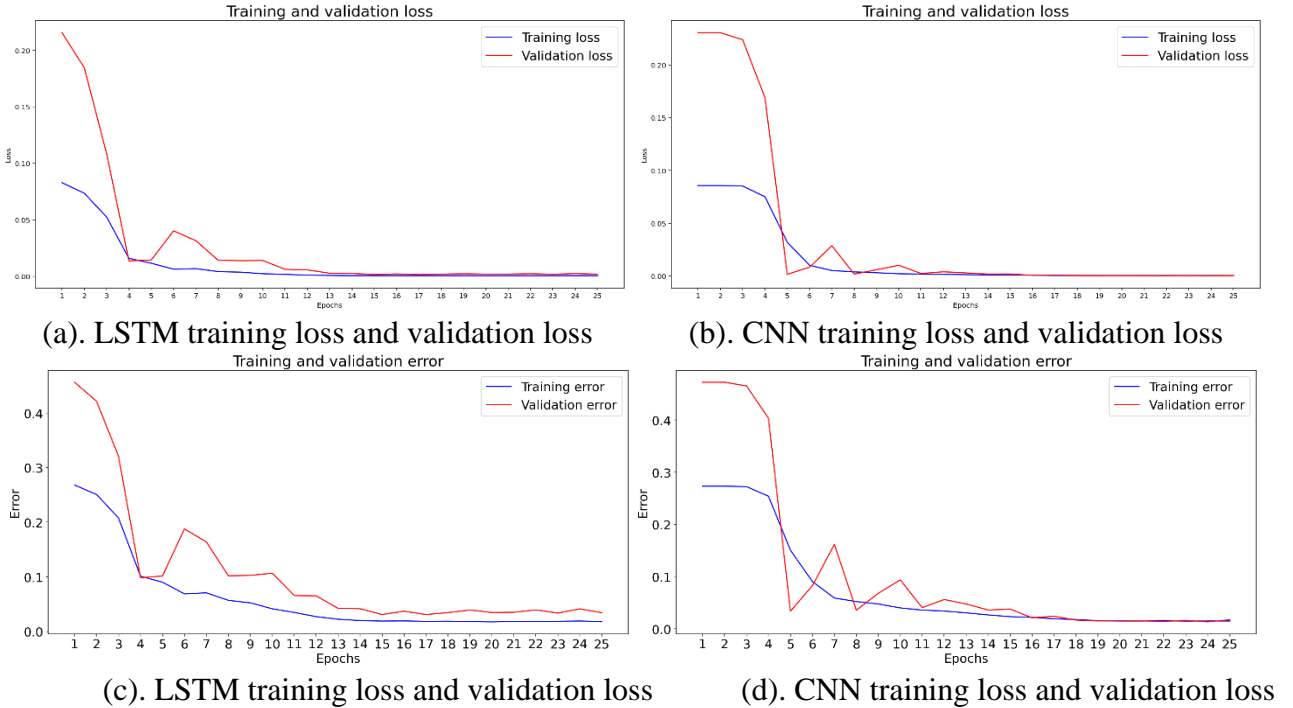
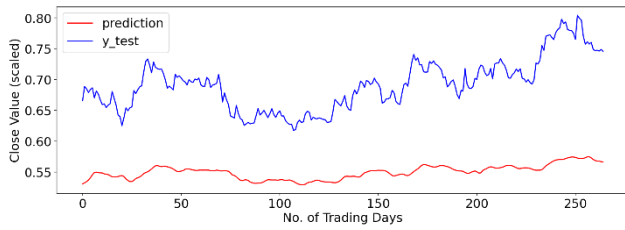


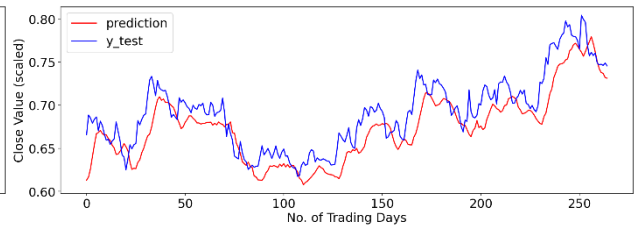
Figure 1: Comparison of LSTM and CNN in loss and error

When the training set and verification set data in the sample are predicted, they are shown in Figure 2. It is obvious that the predicted value of the LSTM model has a large deviation from the real data of the training set and the verification set. Although there are some errors in CNN model, its prediction results are closer to the real situation and its data processing is better.

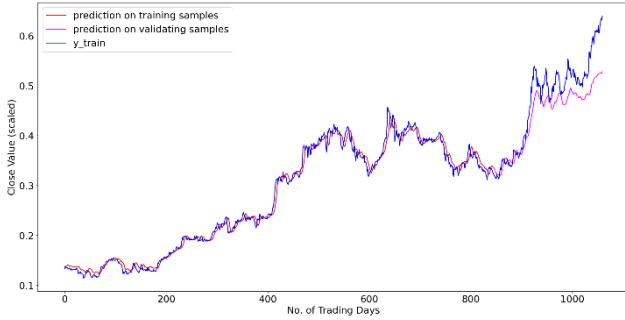
After comparing and analysing the data of the training set and the verification set of the two models, the two models are respectively used to predict the stock price of the time period in which the test set is located. The Figure 3 shows the prediction results of the two models. It is obvious that the predicted value under the CNN model is very close to the real value, and the image fitting effect is better than that of the LSTM model, resulting in smaller errors. Therefore, the effect of CNN in stock price prediction is obviously better than that of LSTM.



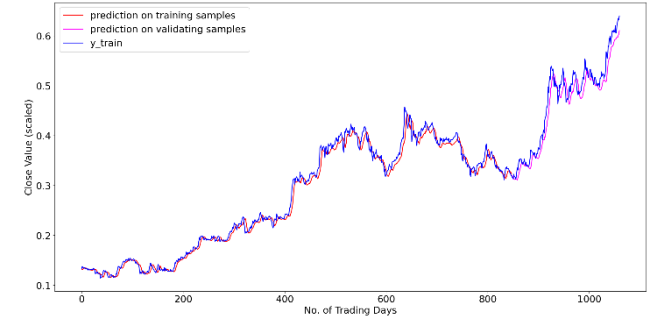
(a). the predicted value and the real value for LSTM



(b). the predicted value and the real value for LSTM



(c). predicted values and real values of LSTM training and validation

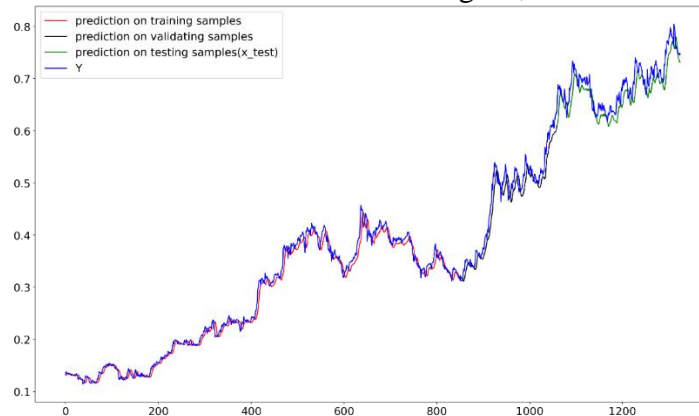


(d). predicted values and real values of CNN training and validation

Figure 2: Comparison of LSTM and CNN in predicted and real values



(a). Predicted and real values of LSTM training set, validation set and test set



(b). Predicted and real values of CNN training set, validation set and test set

Figure 3: Comparison of LSTM and CNN in predicted and real values of training set, validation set and test set.

5. Result

We used RMSE, MAE and MAPE to evaluate the predictive effect of the two models. According to the Table 2, the prediction effect of CNN model is better than that of LSTM model. Through MAE calculation, it can be seen that the MAE value of LSTM is 14135.0985, which is much larger than 2002.2206 of CNN model. In the calculation of MAPE, the percentage error of CNN model is nearly 17.4% lower than that of LSTM. Therefore, the CNN model can predict the closing price well relative to the LSTM model.

Table 2: The values of RMSE, MAE, MAPE comparison between LSTM and CNN model

	Train RMSE	Train MAE	Test RMSE	Test MAE	MAE	MAPE
LSTM	0.0219	0.0133	0.1450	0.1414	14135.0985	20.2513
CNN	0.0140	0.0099	0.0250	0.0200	2002.2206	2.8666

6. Conclusion

This paper compares the prediction of the closing price of stock price by CNN and LSTM models, and takes the opening price, closing price, highest price and lowest price as the input data. The two models are used to train, verify and test the input data respectively to predict the closing price of Madras Rubber Factory (MRF). The experimental results show that compared with LSTM, the MAPE value of CNN model is smaller, and it has higher precision and accuracy in stock prediction.

Therefore, CNN model can be well used in stock price prediction and provide reference information for people in stock trading. However, because the change of stock price is very easy to be affected by the national situation, economic market changes, this model still has some errors in stock price prediction. In the future, more factors can be taken into account to find a more accurate model for stock prediction.

References

- [1] Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47(1062-9408), 552–567. <https://doi.org/10.1016/j.najef.2018.06.013>.
- [2] Li, Y., & Pan, Y. (2021). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, 13(139–149). <https://doi.org/10.1007/s41060-021-00279-9>.
- [3] Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, 48(1). <https://doi.org/10.1029/2010wr009945>.
- [4] Marcus, G. (2018). *Deep Learning: A Critical Appraisal*. ArXiv.org. <https://arxiv.org/abs/1801.00631>.
- [5] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77(0031-3203), 354–377. <https://doi.org/10.1016/j.patcog.2017.10.013>.
- [6] Abdou, M. A. (2022). Literature review: efficient deep neural networks techniques for medical image analysis. *Neural Computing and Applications*, 34(5791–5812). <https://doi.org/10.1007/s00521-022-06960-9>.
- [7] Han, D., Liu, Q., & Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications*, 95(0957-4174), 43–56. <https://doi.org/10.1016/j.eswa.2017.11.028>.
- [8] Han, Y., Huang, G., Song, S., Yang, L., Wang, H., & Wang, Y. (2022). Dynamic Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 7436–7456. <https://doi.org/10.1109/TPAMI.2021.3117837>.
- [9] Hubel, D., & Wiesel, T. (1977). Ferrier lecture - Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 198(1130), 1–59. <https://doi.org/10.1098/rspb.1977.0085>.
- [10] Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017, October 1). Research on data augmentation for image classification

- based on convolution neural networks. *IEEE Xplore*. <https://doi.org/10.1109/CAC.2017.8243510>.
- [11] Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(5455–5516). <https://doi.org/10.1007/s10462-020-09825-6>.
- [12] Wang, S., Li, Z., & Zhao, X. (2022, January 28). *The Application of Convolutional Neural Network in Malware Images Classification*. *Www.atlantis-Press.com*; Atlantis Press. <https://doi.org/10.2991/assehr.k.220110.047>.
- [13] Luo, J.-H., Wu, J., & Lin, W. (2017). *ThiNet: A Filter Level Pruning Method for Deep Neural Network Compression*. *Openaccess.thecvf.com*. https://openaccess.thecvf.com/content_iccv_2017/html/Luo_ThiNet_A_Filter_ICCV_2017_paper.html.
- [14] Deshmukh, A. M. (2020). *Comparison of Hidden Markov Model and Recurrent Neural Network in Automatic Speech Recognition*. *European Journal of Engineering and Technology Research*, 5(8), 958–965. <https://doi.org/10.24018/ejeng.2020.5.8.2077>.
- [15] Pham, V., Bluche, T., Kermorvant, C., & Louradour, J. (2014, September 1). *Dropout Improves Recurrent Neural Networks for Handwriting Recognition*. *IEEE Xplore*. <https://doi.org/10.1109/ICFHR.2014.55>.